
FAMILY CORONAVIRIDAE

TAXONOMIC STRUCTURE OF THE FAMILY

ORDER	<i>Nidovirales</i>
FAMILY	<i>Coronaviridae</i>
GENUS	<i>Coronavirus</i>
GENUS	<i>Torovirus</i>

GENUS CORONAVIRUS

Type Species *Infectious bronchitis virus (IBV)*

VIRION PROPERTIES

Morphology

Virions are enveloped and spherical in shape; those of *Coronavirus* being commonly 120–160 nm in diameter, with an internal, possibly icosahedral, core shell of around 65 nm, and a helical nucleocapsid (Fig. 1). Coronaviruses have large surface projections formed by glycoproteins (peplomers) with a globular and a stem portion. The peplomers (trimers of the spike protein) are about 20 nm in length. In some coronaviruses such as *Bovine coronavirus (BCoV)* and *Murine hepatitis virus (MHV)*, a second layer of peplomers formed by the hemagglutinin esterase (HE) protein is also observed. A gap separating the internal core from the envelope has been observed in coronaviruses using cryoelectron microscopy. The core can be released after treatment with detergents. Disruption of these cores releases N-protein-containing helical nucleocapsids.

Physicochemical and Physical Properties

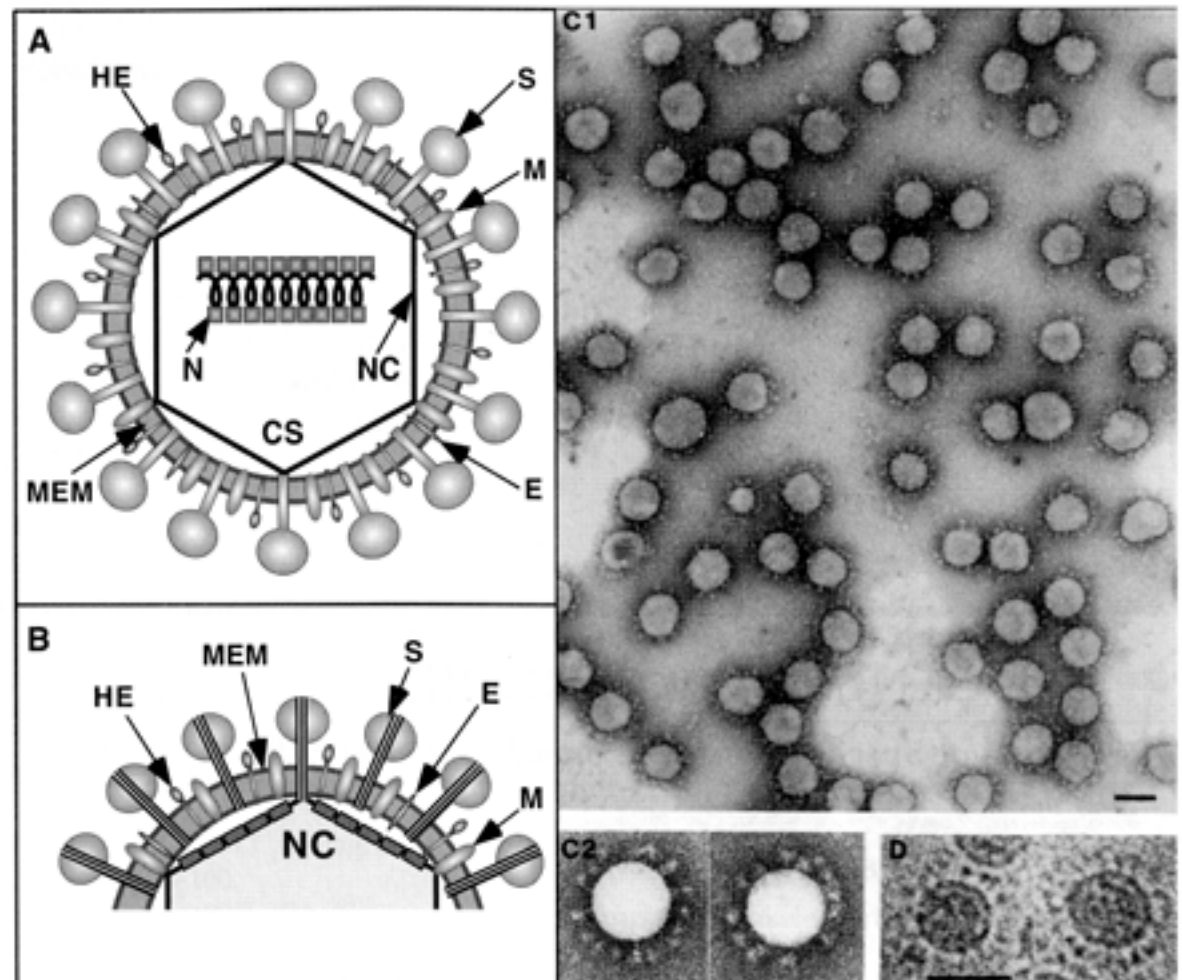
Virion Mr has been estimated at 400×10^6 for coronaviruses. Virion buoyant density in sucrose is 1.15–1.20 g/cm³; density in CsCl is 1.23–1.24 g/cm³. Virion S_{20w} is 300–500S. Virions are sensitive to heat, lipid solvents, nonionic detergents, formaldehyde, oxidizing agents, and UV irradiation. After incubating for 24 hours at 37°C a 10-fold or more decrease in virus infectivity has been observed with certain strains in tissue culture. Magnesium ions (1 M) reduce the extent of heat inactivation in MHV. Some viruses in both genera are stable at pH 3.0.

Nucleic Acid

Virions contain a single molecule of linear, positive sense, ssRNA. The genomic RNA is the largest viral RNA genome known ranging from 27.6 to 31 kb (*Coronavirus*)

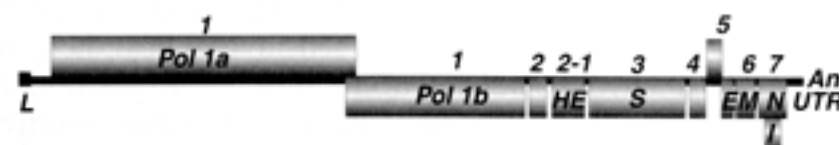
FIGURE 1

Structure of coronavirus virions. (A) Schematic diagram of virus structure. (B) Diagram of virion surface. (C1) Electron micrograph of virus particles of *Transmissible gastroenteritis virus* (TGEV) stained with uranyl acetate (C1) or sodium phosphotungstate (C2) showing the surface of the virus particles. The peplomers are better defined using sodium phosphotungstate. (D) Cryoelectron microscopic visualization of unstained TGEV in vitreous ice. The particles contain an internal structure inside the viral envelope and well-extended peplomers. MEM, lipid membrane; S, spike protein; M, large membrane protein, E, small envelope protein; HE, hemagglutinin-esterase; N, nucleocapsid protein; CS, core shell; NC, nucleocapsid. The bars represents 100 nm.



in size. Coronavirus RNA has a 5'-terminal cap followed by a leader sequence of 65–98 nts and an untranslated region of 200–400 nts. At the 3' end of the genome is an untranslated region of 200–500 nts followed by a poly(A) tail. The virion RNA functions as an mRNA and is infectious. It contains approximately 7–10 functional genes, 4 or 5 of which encode structural proteins. The genes are arranged in the order 5'-polymerase-(HE)-S-E-M-N-3', with a variable number of other genes that are believed to be nonstructural (NS) and largely nonessential, at least in tissue culture. A genome structure of a Coronavirus is shown in Fig. 2. The complete

Coronavirus

**FIGURE 2**

Representation of prototype Coronavirus genome. ORFs are represented by boxes. Numbers above boxes indicate the nRNA designation. The pseudoknots in ORF1 are aligned. The proteins encoded by the ORFs are indicated. The 5' leader sequence is depicted by a small black box; poly(A) tail is indicated by An; M, membrane protein; N, nucleocapsid protein; S, spike protein; HE, hemagglutinin-esterase protein; L, internal ORF; empty boxes, nonstructural proteins; L, leader; UTR, untranslated region.

sequence of several coronavirus genomes have been determined and are accessible from international databases (MHV; TGEV; IBV; and *Human coronavirus 229E*, HCoV-229E).

Proteins

Virions contain a large surface glycoprotein (or spike, S), an integral membrane protein (M) which has integrated in the virus envelope three or four segments, a small membrane protein (E) and a nucleocapsid protein (N) (Table 1). The ratios of S/E/M/N proteins vary in different reports. For TGEV, these ratios have been estimated to be 20:1:300:100, respectively. The S protein is large, ranging from 1160 to 1452 amino acids, and in some coronaviruses is cleaved into S1 and S2 subunits. The S protein is responsible for attachment to cells, hemagglutination, membrane fusion, and induction of neutralizing antibodies. Immunization with S alone can induce protection from challenge with some coronaviruses (MHV, TGEV). The S protein has a carboxy terminal half with a coiled-coil structure. The M protein contains 225–260 amino acids and can induce α -interferon. A subset of coronaviruses contains a hemagglutinin-esterase protein (HE) which forms short surface projections. This apparently non-essential protein has a receptor binding domain for 9-O-acetylated neuraminic acid, hemagglutination activity, as well as receptor destroying activities (neuraminase-O-acetylase). The HE protein shows some nucleotide sequence identity with the hemagglutinin-esterase protein of *Influenza C virus*. The E protein (80–109 amino acids), together with the M protein, plays an essential role in coronavirus particle assembly. The N protein (377–455 amino acids) is a highly basic phosphoprotein that modulates viral RNA synthesis, binds to the viral RNA, and forms a helical nucleocapsid.

The nonstructural proteins are generally not essential for virus replication in tissue culture or *in vivo*. One indispensable non-structural protein is the replicase encoded by gene 1, which accounts for two-thirds of the genome (18–22 kb). The replicase gene is predicted to encode a protein of approximately Mr 740–800 $\times 10^3$ which is co-translationally processed. The *pol* gene encodes two ORFs, 1a and 1b, which overlap by a few nts. ORF1b is in the -1 reading frame with respect to ORF1a. Several domains within *pol* have predicted functions based on regions of nucleotide homology including: two papain-like cysteine proteases, a chymotrypsin-picornaviral 3C-like protease, a cysteine-rich growth factor-related protein, an RNA-dependent RNA polymerase, a nucleoside triphosphate (NTP) binding/helicase domain, and a zinc-finger nucleic acid binding domain.

The other nonstructural proteins vary in name and location among coronaviruses. The locations of the genes encoding these nonstructural proteins are indicated in

TABLE 1 Virus-associated proteins of coronaviruses

Protein		Coronavirus ^a
Spike glycoprotein	S	180–220
Membrane protein	M	23–35
Nucleocapsid protein	N	50–60
Small-envelope protein	E	9–12
Hemagglutinin-esterase protein	HE	65

^aApparent Mr estimated by electrophoresis ($\times 10^3$).

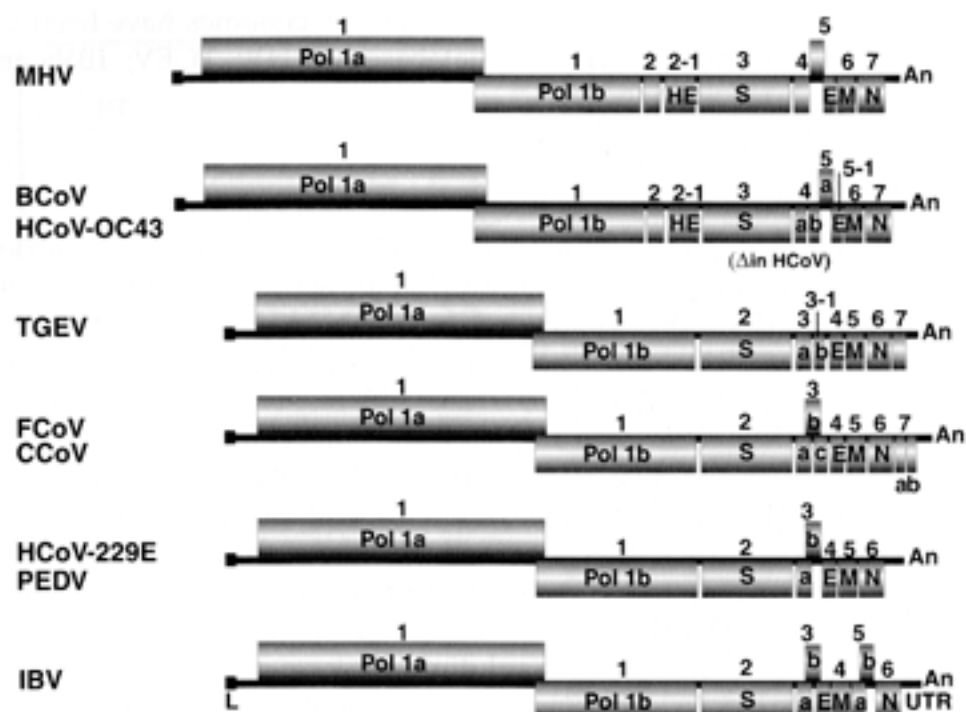


FIGURE 3

Genetic structure in coronaviruses. The numbers above the bars indicate mRNA. The name in the bars indicate the protein encoded by the corresponding ORF. The structural protein genes are marked by various symbols while the nonstructural protein genes are represented by unfilled boxes or by boxes with small letters. The *Human coronavirus OC43* (HCoV-OC43) does not have the two nonstructural proteins encoded by gene 4. The translation products of ORFs *Murine hepatitis virus* (MHV) 5a, *Bovine coronavirus* (BCoV) 4a,b, *Feline coronavirus* (FCoV) and *Canine coronavirus* (CCoV) 3b have not been detected. The ORFs *Porcine transmissible gastroenteritis virus* (TGEV) 3b, FCoV and CCoV 3c, and HCoV-229E 3a,b are homologous. The 3a ORFs of TGEV, FCoV, and CCoV are homologous. The ORFs TGEV 7 and the 7a of FCoV and CCoV are homologous. The ORF abbreviations are as in Fig. 2. L, leader; Δ, deletion; UTR, untranslated region.

Fig. 3. The N gene is usually located at the 3' end in the genome of coronavirus with the exception of TGEV, *Feline coronavirus* (FCoV), and *Canine coronavirus* (CCoV) in which the N gene is followed by one or two other genes.

Lipids

Virions have lipid-containing envelopes derived from the host cell. The S protein of coronaviruses is acylated (MHV, BCoV). The MHV E protein is also acylated.

Carbohydrates

The S and HE proteins contain N-linked glycans and the S protein is extensively glycosylated. The M protein of coronaviruses contains a small number of either N- or O-linked glycans, depending on the strain.

GENOME ORGANIZATION AND REPLICATION

About two-thirds of the entire RNA is occupied by the polymerase gene. At the overlap between the ORF1a and 1b regions, there is a specific seven-nucleotide "slippery" sequence and a pseudoknot structure (ribosomal frameshifting signi-

which are required for the translation of ORF1b. The 3' one-third of the genome comprises the genes encoding the structural proteins and the other nonstructural ones. The organization of the nonstructural protein genes, which are interspersed between the known structural protein genes, varies significantly among different coronaviruses (Fig. 3). A pseudoknot structure is also predicted at the 3' end of the coronaviral RNA.

Coronavirus RNA synthesis occurs via an RNA-dependent RNA synthesis process in which mRNAs are transcribed from negative-stranded templates. There is a stretch of the consensus sequence UCUAAAC (for MHV), or a highly related sequence for other coronaviruses, at sites immediately upstream of most of the genes. These sequences represent signals for the transcription of sgRNAs. Coronavirus mRNAs consist of six to eight types of varying sizes, depending on the coronavirus and the host species. The largest mRNA is the genomic RNA which also serves as the mRNA for ORF1a and 1b; the remainder are sgRNAs. The RNAs are designated mRNA 1–7, in order of decreasing size. Some mRNAs have been given a hyphenated name, e.g., mRNA-2-1, because they were discovered after the original set of described mRNAs. The mRNAs have a nested-set structure in relationship to the genome structure (Fig. 4). Except for the smallest mRNA, all of the mRNAs are structurally polycistronic. In general, only the 5'-ORF of each mRNA is translated. However, there are exceptions: some mRNAs, e.g., mRNA-5 of MHV, mRNA-3 of IBV, and BCoV nucleocapsid mRNA are translated by internal initiation into two or three proteins.

Coronavirus mRNAs have another unique structural feature: their 5' ends have a leader sequence of approximately 65–98 nts, which is derived from the 5' end of the genomic RNA. At the mRNA start sites on the viral genomic RNA, there is a short stretch of sequence that is nearly homologous to the 3' end of the leader RNA. This sequence constitutes part of the signal for sgRNA transcription.

Coronavirus RNA synthesis occurs in the cytoplasm via a negative strand RNA intermediate. Both genome size and subgenomic negative strand RNAs, which correspond in number of species and size to those of the virus-specific mRNAs have been detected. The 5' end of the negative strand RNA contains short stretches of oligo(U). The subgenomic negative strand RNAs appear to be mirror images of the positive strand subgenomic RNAs. Several transcription models have been proposed, each of which is consistent with some of the experimental data. These models are not mutually exclusive, as components of each model may operate at different stages of the viral replication cycle. The two models compatible with most of the experimental data are leader-primed transcription and discontinuous

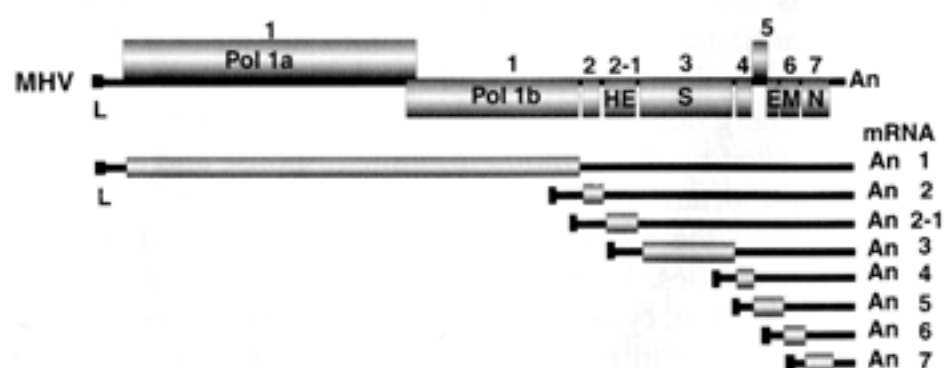


FIGURE 4

Structural relationship between mRNAs and the genomic RNA of coronaviruses. Thick lines represent the translated sequences; thinner lines, untranslated sequences. Numbers above the boxes indicate the mRNA designation. The names in the boxes indicate the proteins encoded by the corresponding genes. L, leader. Abbreviations are as in Fig. 2.

transcription during negative strand RNA synthesis. The *leader-primed transcription* model proposes that the virion genomic RNA is first transcribed into a genomic length, negative strand RNA, which, in turn, becomes the template for subsequent sgRNA synthesis. The leader is transcribed from the 3' end on the negative-strand RNA and dissociates from the template to subsequently associate with the template RNA at the various mRNA start sites serving as a primer for the transcription of the viral sgRNAs. It is proposed that the discontinuous transcription step takes place during positive strand RNA synthesis. In contrast to the leader-primed transcription model, the *discontinuous transcription during negative strand RNA synthesis* model proposes that the discontinuous transcription step occurs during negative strand RNA synthesis, generating subgenomic negative strand RNAs, which then serve as templates for sgRNAs in interrupted transcription. In this model, the intergenic sequences on the genomic RNA at the mRNA start sites serve as termination or pausing signals for negative-strand synthesis, and the nascent subgenomic negative strand RNA then jumps to the leader RNA sequence at the 5' end of the genomic RNA to act as a primer for transcription.

The packaging signal for MHV RNA, as determined using defective minigenomes, is localized near the 3' end of gene 1. This packaging signal forms a stem loop and is sufficient for the packaging of DI RNA or heterologous RNAs into virions.

Coronaviruses undergo recombination at very high frequency during the replication of positive and negative strand RNAs. This is particularly true for MHV. A lower recombination frequency has been described for IBV and TGEV.

The assembly of virus particles probably starts with the formation of a ribonucleoprotein (RNP) which interacts with components of the core shell. Virions mature in the cytoplasm by budding through the endoplasmic reticulum and other pre-Golgi membranes. The interaction between the M and E proteins appears to be a key event for virus particle assembly. The S and HE proteins are not necessary for virus particle formation.

ANTIGENIC PROPERTIES

Strong humoral immune responses are elicited by the four structural proteins (S, M, and N, and HE protein when present). The S and HE proteins are the predominant antigens involved in virus neutralization. Reduction of infectivity with anti-M antibodies also has been shown, but generally in the presence of complement. Protection against *Coronavirus* infections (MHV, TGEV) is provided by S protein that is affinity-purified or expressed by recombinant adenovirus. N- and M-specific antibodies also give some protection *in vivo*. The most efficient induction of virus neutralizing antibodies has been achieved with a combination of S and N proteins. The globular portion of the S protein contains many dominant antigenic sites targeted by the humoral immune response and also by cytotoxic T lymphocytes. Other important antibody epitopes are also found in the stem portion, at least for MHV. Both the amino and the carboxy termini of the M protein elicit strong immune responses. While neutralizing antibodies can prevent disease if present prior to infection, cytotoxic T cell responses are important in virus clearance. Hypervariable domains in the S1 portion of the S protein facilitate the selection of virus escape mutants that evade both humoral and cellular immune responses. N protein also elicits a protective cellular immune response.

The immune system is known to play a major role in the pathogenesis of coronavirus-induced disease from the experimental induction of demyelination and the antibody-dependent enhancement of FCoV infectivity.

BIOLOGICAL PROPERTIES

Coronaviruses infect birds and many mammals, including humans. The respiratory track, gastrointestinal organs, and neurological tissues are the most common targets of coronaviruses, but other organs including liver, kidney, heart, and eye can also be affected. Epithelial cells are the main target of coronaviruses. Widely distributed cells such as macrophages are also infected by coronaviruses. These viruses have relatively restricted host ranges, infecting only their natural host and closely related animal species. Occasionally, cross-species infection by coronaviruses occurs, such as the experimental infection of dogs by TGEV. Biological vectors are not known. Respiratory, fecal-oral, and mechanical transmission are common.

Although coronaviruses may bind cells through ubiquitous acetylated forms of glycoproteins and lipids, a more specific binding between the virus and a cellular receptor is required for the establishment of viral infection. Coronaviruses are divided into three groups (see below). Coronavirus of group 2, including MHV (see below), use as receptors members of the biliary glycoprotein (bgp) subfamily of the carcinoembryonic antigen (CEA) family. Members of group 1 of coronaviruses (including TGEV and HCoV-229E; see below) use the aminopeptidase N (APN or CD13) as a receptor for cell entry. Sialic acid (N-acetyl-9-O-acetylneuraminic acid)-containing glycoproteins are probably a component of the cell surface molecules required for BCoV and HCoV-OC43 infectivity. Nevertheless, binding to bgp and APN receptors is not sufficient for viral infection and does not explain the differences in tropism of coronaviruses. In addition to the binding to the afore-mentioned receptors, a second factor mapping in the S protein (possibly, a second receptor binding site) has been implicated in *Coronavirus* tropism.

Infections of humans and animals by coronaviruses seem to be ubiquitous, as evidence of infection has been obtained in every country where serological or virological studies have been done.

LIST OF SPECIES DEMARCATION CRITERIA IN THE GENUS

In the genus *Coronavirus* the differentiation among species is essentially based on:

- the organization and sequence of the nonstructural proteins in the species genome,
- antigenic properties,
- the processing of the S protein into two halves (S1 and S2), and
- the host.

The number and location of the nonessential genes (frequently encoding nonstructural proteins) significantly varies in composition and location. These nonessential genes are located either between genes: Pol and S, genes S and M, genes M and N, or after gene N (Fig. 3; for details, see Lai and Cavanagh, 1997). For instance, MHV and BCoV of coronavirus group 2 may have ORFs 2 and 3-1 encoding a protein of 260 amino acids and the HE, respectively, which are absent in coronaviruses of groups 1 and 3. The presence of ORFs between genes M and N is unique to the IBV species, and only the canine and feline viruses have two ORFs after gene N. By contrast, the N gene is the last one identified in MHV, BCoV, HCoV-229E, PEDV, and in IBV (Fig. 3).

Serological characterization of group 1 coronaviruses has shown that there are three distinct antigenic clusters, one formed by TGEV, CCoV, and FCoV, and two others formed by HCoV-229E and PEDV.

The processing of S protein into S1 and S2 halves differentiates species of group 1 (TGEV, CCoV, FCoV, HCoV-229E and PEDV), in which S protein is not cleaved from species of groups 2 (MHV, BCoV, and HCoV-OC43) and 3 (IBV), in which the S protein is cleaved.

LIST OF SPECIES IN THE GENUS

Official virus species names are in italics. Tentative virus species names, alternative names (), strains, or serotypes are not italicized. Virus names, genome sequence accession numbers [], and assigned abbreviations () are:

Species in the Genus

Group 1 species

<i>Canine coronavirus</i>	[DL3096]	(CCoV)
<i>Feline coronavirus</i>		(FCoV)
<i>Feline infectious peritonitis virus</i>		(FIPV)
<i>Human coronavirus 229E</i>	[X69721]	(HCoV-229E)
<i>Porcine epidemic diarrhea virus</i>	[Z35758]	(PEDV)
<i>Transmissible gastroenteritis virus</i>	[Z24675, Z34093,	(TGEV)
<i>Porcine respiratory coronavirus,</i>	D00118, X06371]	(PRCoV)

Group 2 species

<i>Bovine coronavirus</i>		(BCoV)
<i>Human coronavirus OC43</i>		(HCoV-OC43)
<i>Murine hepatitis virus</i>	[AF029248]	(MHV)
<i>Porcine hemagglutinating encephalomyelitis virus</i>		(HEV)
<i>Rat coronavirus</i>		(RtCoV)
<i>Sialodacryoadenitis virus,</i>		(SDAV)

Group 3 species

<i>Infectious bronchitis virus</i>	[M95169]	(IBV)
<i>Turkey coronavirus</i>		(TCoV)

Tentative Species in the Genus

<i>Rabbit coronavirus</i>		(RbCoV)
---------------------------	--	---------

GENUS *TOROVIRUS*

Type Species *Equine torovirus* (EqTV)

DISTINGUISHING FEATURES

The nucleocapsid has a tubular appearance and virions are disk-, kidney-, or rod-shaped (Fig. 5). Data are based mostly on the *Equine torovirus* (EqTV) strain Berne and on *Bovine torovirus* (BoTV) strain Breda. There are four structural proteins: N, M, S, and HE. The HE protein is dispensable for replication *in vitro*; in EqTV,